



# UNITED STATES PATENT AND TRADEMARK OFFICE

*MN*  
UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/823,685	04/14/2004	Randall J. Calistri-Yeh	55653-017	5370
<div>7590 McDermott, Will &amp; Emery 600 13th Street, N.W. Washington, DC 20005-3096</div>			<div>EXAMINER BUSS, BENJAMIN J</div>	
			<div>ART UNIT 2129</div>	<div>PAPER NUMBER</div>
			<div>MAIL DATE 05/16/2007</div>	<div>DELIVERY MODE PAPER</div>

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

## Office Action Summary

Application No.

10/823,685

Applicant(s)

CALISTRI-YEH ET AL.

Examiner

Benjamin Buss

Art Unit

2129

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

### Status

- 1) ☒ Responsive to communication(s) filed on 21 March 2007.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

### Disposition of Claims

- 4) ☒ Claim(s) 53-59, 65 and 71 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 53-59, 65 and 71 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 4/14/2004 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

### Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
  - ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.

### Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)  
Paper No(s)/Mail Date 4/14/04 & 3/21/07.
- 4) ☐ Interview Summary (PTO-413)  
Paper No(s)/Mail Date. \_\_\_\_\_.
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other: \_\_\_\_\_.

**DETAILED ACTION**

This Office Action is in response to an AMENDMENT entered 8/18/2006 for the patent application 10/823,685 filed on 4/14/2004, which claims priority based on 09/562,916 filed on 5/2/2000 and 60/177,654 filed on **1/27/2000**. The Office Actions of 11/2/2006 and 4/13/2006 are fully incorporated into this Office Action by reference. Claims 53-59, 65, & 71 are pending.

In the event that Applicant chooses to amend, Examiner suggests further defining these broad terms in the claims:

- **relative strength**
- **degree of relevance**

**Information Disclosure Statement**

The Information Disclosure Statements of 4/14/2004 and 3/21/2007 have been considered and placed in the Application file.

**Specification**

The specification is objected to as failing to provide proper antecedent basis for the claimed subject matter. See 37 CFR 1.75(d)(1) and MPEP § 608.01(o). Correction of the following is required:

- Claim 71 recites a "tangible computer-readable medium". This is not supported by the specification. The claim must be amended such that it is supported by the disclosure as originally filed.

Appropriate correction is required.

**Claim Objections**

Claims 53, 57, 65, and 71 are objected to because of the following informalities:

- Claim 53 L24: Change "the predetermined particular categories" to -- the predetermined categories --.
- Claim 53 L25: Change "relative each of the predetermined particular categories" to -- relative to each of the predetermined categories --.
- Claim 57 L19: Change "the predetermined particular categories" to -- the predetermined categories --.

Art Unit: 2129

- Claim 57 L20: Change "relative each of the predetermined particular categories" to -- relative to each of the predetermined categories --.
- Claim 65 L25: Change "the predetermined particular categories" to -- the predetermined categories --.
- Claim 65 L26: Change "relative each of the predetermined particular categories" to -- relative to each of the predetermined categories --.
- Claim 71 L25: Change "the predetermined particular categories" to -- the predetermined categories --.
- Claim 71 L26: Change "relative each of the predetermined particular categories" to -- relative to each of the predetermined categories --.

Appropriate correction is required.

**Claim Rejections - 35 USC § 112**

The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claim 71 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

- Claim 71 claims a "tangible computer-readable medium", but it is not clear what the metes and bounds of "tangible computer-readable medium" are.

Appropriate correction is required.

**Response to Arguments**

Applicant's arguments, see page 9, filed 3/21/2007, with respect to claim 71 have been fully considered and are persuasive. The rejection of claim 71 under 35 U.S.C. §112, second paragraph, is based on a new grounds necessitated amendment.

**Claim Rejections - 35 USC § 101**

35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Art Unit: 2129

Claims 53-59, 65, and 71 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter. The computer system must set forth a practical application of a §101 judicial exception to produce a real-world result. Benson, 409 U.S. at 71-72, 175 USPQ at 676-77. The invention is ineligible because it has not been limited to a substantial practical application. An invention directed to selecting abstract datasets or data points is useless in a real world situation. Further adding of selected abstract data points to an abstract dataset is also useless in a real world situation. Additionally generating some abstract result based on the abstract manipulations is also useless in a real world situation.

"[T]aking several abstract ideas and manipulating them together adds nothing to the basic equation." AT&T v. Excel at 1453 quoting In re Warmerdam, 33 F.3d 1354, 1360 (Fed. Cir. 1994).

In determining whether the claim is for a "practical application," the focus is not on whether the steps taken to achieve a particular result are useful, tangible, and concrete. If the claim is directed to a practical application of the §101 judicial exceptions producing a **result tied to the physical world** that does not preempt the judicial exception, then the claim meets the statutory requirement of 35 U.S.C. §101.

The phrases 'constructing a semantic vector', 'receiving a query containing information', 'comparing the semantic vector for the query to the semantic vector of each dataset', 'selecting datasets whose semantic vectors are closest in distance to the semantic vector for the query', 'outputting information', 'relative strength', 'degree of relevance', 'represents the relative strength', 'using a method other than semantic vectors' are not clear in purpose or scope. Other variations on these phrases in the claims do not provide a clear purpose or scope for the claimed invention.

The invention must be for a practical application and either:

- 1) specify transforming (physical thing – article) or
  - 2) have the Final Result (not the steps) achieve or produce a
    - useful (specific, substantial, AND credible),
    - concrete (substantially repeatable/non-unpredictable), AND
    - tangible (real world/non-abstract) result
- (tangibility is the opposite of abstractness).

Art Unit: 2129

A claim that is so broad that it reads on both statutory and non-statutory subject matter must be amended, and if the specification discloses a practical application but the claim is broader than the disclosure such that it does not require the practical application, then the claim must be amended.

5 Claims that construct data representations, receive queries, construct further data representations, compare data representations, select data based on similarities between abstract representations, generate an abstract result, and combining abstract data are not statutory. In more detail to clarify the issues:

Applicant has provided examples of how to determine the semantic vectors that are "closest in distance" to the semantic vector for the query (Specification, ¶[0073] & ¶[0105] of the PG-PUB), but no claim limitations are directed to the method of measuring distance actually used by the invention. Examiner finds that the "selecting  
10 datasets whose semantic vectors are closest to the semantic vector for the query" is not concrete because it is not clear what measure of closeness Applicant uses. As described and claimed, the "closest in distance" vector could be determined based on Hamming distance, eyeball guesstimation, Manhattan distance, human intuition, Chebyshev distance, with a tape measure, etc. Distance can be measured in many non-repeatable ways.

Applicant has provided examples of how to determine the "significance of each data point" (Specification,  
15 ¶[0051]-[0056] of the PG-PUB), but no claim limitations are directed to the method of determining a relative strength or a degree of relevance actually used by the invention to determine the significance of each data point. Examiner finds that the "determining the significance of each data point with respect to the predetermined categories" is not concrete because it is not clear what measure of "relative strength" or "degree of relevance". As described and claimed, the "significance" could be determined based on any arbitrary determination of strength or relevance.

20 The result of claims 53-56, 65, & 71 is "outputting information of the selected datasets to be corresponding to the desired datasets identified in the query". This result is abstract because it is not clear what "outputting information" includes. It is not necessarily relied upon for a practical application. This result is not concrete at least because the "significance" and measure of "closest" are not concrete. This result is not tangible at least because it is merely an abstract result generated by an abstract manipulation of abstract data that has no grounding in the real world. This result is not useful because Applicant has not provided a real world practical application which relies  
25 upon the generated result. The claimed result is not tied to the physical world.

The result of claims 57-59 is merely "associating said selected data points to said dataset". This "association" is reasonably interpreted to be an abstract manipulation. Applicant has not shown that the resulting

Art Unit: 2129

“association” is relied upon for a practical application. The resulting claimed “association” is not tied to the physical world. This resulting claimed “dataset” is not concrete at least because all of the “datasets”, the “data points” the “significance”, and the measure of “closest” are not concrete. This result is not tangible at least because it is merely an abstract result generated by an abstract manipulation of abstract data that has no grounding in the real world. This result is not useful because Applicant has not provided a real world practical application which relies upon the generated result.

Furthermore, for claims 53, 55-57, 65, and 71: Applicant has claimed that the dataset(s) include “one or more data points and each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string”. Examiner finds that the “datasets” and “data points” are not concrete because it is not clear what they are. As claimed, the datasets are reasonably interpreted to include at least any of the following abstract datasets:

- character strings representing mathematical equations;
- images of arbitrary fractals, which are approximately stochastically similar geometric shapes;
- text for a computer program;
- any arbitrary conglomeration of symbols;
- a dramatic pause punctuating a speech;
- etc.

The manipulation of abstract data is not statutory.

Claim 71 is further rejected as non-statutory for being directed to a “tangible computer-readable medium carrying one or more sequences of instructions” where the specification provides for the computer-readable medium including mediums such as: punch cards, paper tape, any other physical media with patterns of holes, and transmission media such as coaxial cables, copper wire, fiber optics, acoustic or light waves, and carrier waves. Please recognize that for claims where applicant is seeking to patent *functional descriptive material in combination with some media*, there are currently three requirements that need to be met, or a rejection is warranted. 1) The media needs to only cover (i.e., be limited to) embodiments which establish a statutory category of invention. The current Office position is that signals, **waves**, radiation and other such media are not physical articles or objects and as such are not manufactures or machines under 35 U.S.C. §101. Since they are clearly not processes or compositions, they fail to fall within a statutory category. 2) The media must be structurally and functionally

Art Unit: 2129

interconnected with the functional descriptive material in such a manner that it enables the functional descriptive material to act as a computer component and realize its functionality. The current office position is that signals, **waves**, radiation, wires, fibers, and **printed matter** in and of themselves fail to meet this criteria. As such, no usefulness can be gleaned from the functional descriptive material. 3) The functional descriptive material must

5 provide a practical application of the idea embodied therein when executed as a computer component.

Merely adding pre-pending "tangible" to the "computer-readable medium" terminology fails remedy the issue because it is not clear what the metes and bounds of "tangible computer readable medium" are. Unless Applicant can show that "tangible computer-readable medium" is present and described or defined in the original disclosure, it is not acceptable in the claims.

10 The Examiner hopes that this additional detail helps Applicant in determining what claim amendments are necessary. Appropriate corrections are required.

#### ***Response to Arguments***

Applicant's arguments filed 3/21/2007 have been fully considered but they are not persuasive. In re page 10, Applicant argues that the amendment to "closest in distance" clarifies that the data sets are selected based on the distance between semantic vectors. Examiner disagrees that this overcomes the outstanding rejection under 35

15 U.S.C. §101. The specification (e.g. ¶¶[0073],[0105]) still leaves this distance measurement open to un-repeatable distance measuring methods such as eyeball guesstimation, human intuition, with a tape measure by declaring "any typical distance measure" acceptable. On this basis, Examiner finds Applicant's argument to be unpersuasive and the rejections STAND.

20 In re pages 10-11, Applicant argues that "significance of each data point" is now clearly defined in the claims. Examiner disagrees. The definition Applicant gives is based on "relative strength" or "degree of relevance". The meaning or definition of these phrases is not clear. What is a "relative strength"? What is a "degree of relevance"? On this basis, Examiner finds Applicant's argument to be unpersuasive and the rejections STAND.

In re page 11, Applicant argues that "result" has been deleted from the claims. Examiner disagrees that this

25 overcomes the outstanding rejection under 35 U.S.C. §101. A useful, concrete, and tangible "final result" is not found in the claims. On this basis, Examiner finds Applicant's argument to be unpersuasive and the rejections STAND.

In re pages 11-12, Applicant argues that data points being "at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string" are physical world examples tied to the real



Art Unit: 2129

world. Applicant further argues that processing data related to data points, datasets, and queries to generate representative semantic vectors thereof, to differentiate between related and non-related datasets produces a useful, non-abstract result analogous to the patentable subject matter of AT&T Corp. v. Excel Communications, Inc.

Examiner disagrees. The data points are not necessarily tied to the physical real world. By way of example, not all  
5 of the following are tied to the physical world: a character string representing mathematical equations, images of arbitrary fractals, text for a computer program, and arbitrary conglomerations of symbols, a dramatic pause punctuating a speech. Since the data points are reasonably interpreted to be abstract information, the result after abstract/mathematical manipulation is also abstract, and therefore not analogous to the patentable subject matter of AT&T Corp. v. Excel Communications, Inc. From that case: "[T]aking several abstract ideas and manipulating them  
10 together adds nothing to the basic equation." AT&T v. Excel at 1453 quoting In re Warmerdam, 33 F.3d 1354, 1360 (Fed. Cir. 1994).

This holding by the Federal Circuit logically follows from the holding in Diehr that quotes Deener. If a process requires "certain things" to be done with "certain substances" in a "certain order," then a claim for doing "certain things" with non-"substances" (e.g., abstract ideas) in a "certain order" fails the Supreme Court standard...there is no  
15 "substance being acted upon...only abstract ideas. The Federal Circuit used this logic to say that: "taking several abstract ideas and manipulating them together adds nothing to the basic equation."

The Federal Circuit held that Warmerdam was dispositive because the Supreme Court finds that issue to be dispositive. The Warmerdam standard is merely an application of the Supreme Court Standard from Diamond v. Diehr and Cochrane v. Deener.

20 Examiner declines to swim against this current, no matter how much Applicant wishes otherwise. Examiner's rejection is merely an application of the same standard that was used by all the cases from Cochrane v. Deener to Warmerdam.

Applicant's claims for mere manipulation of "datasets" including "at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string" thereby fail this standard because  
25 the "datasets" represent mere abstract ideas...mere information in the abstract.

Examiner reads the claims as a whole to carefully search for claim limitations to practical applications and finds none. It is Examiner's opinion that the claims are devoid of statutory claim limitations.

Art Unit: 2129

Having been given ample opportunity to respond by amendment, Applicant has presented no other statutory claim limitations to circumscribe the metes and bounds of the claims sufficiently to change this assessment.

Accordingly, Applicant has failed to carry his burden of showing how the claims are in any way statutory. On this basis, Examiner finds Applicant's argument to be unpersuasive and the rejections STAND.

5           In re page 12, Applicant argues that amending claim 71 to specify "tangible computer-readable medium" addresses the Examiner's concern. Examiner disagrees. Merely adding pre-pending "tangible" to the "computer-readable medium" terminology fails remedy the issue because it is not clear what the metes and bounds of "tangible computer readable medium" are. Unless Applicant can show that "tangible computer-readable medium" is present and described or defined in the original disclosure, it is not acceptable in the claims.

10           Computer programs embodied in a tangible medium, such as floppy diskettes, are patentable subject matter under 35 U.S.C. §101. However, the specification of the instant application describes the computer-readable medium as including mediums such as: punch cards, paper tape, any other physical media with patterns of holes, and transmission media such as coaxial cables, copper wire, fiber optics, acoustic or light waves, and carrier waves. Please recognize that for claims where Applicant is seeking to patent *functional descriptive material in combination*  
15 *with some media*, there are currently three requirements that need to be met, or a rejection is warranted. 1) The media needs to only cover (i.e., *be limited to*) embodiments which establish a statutory category of invention. The current Office position is that signals, **waves**, radiation and other such media are **not** physical articles or objects and as such are not manufactures or machines under 35 U.S.C. §101. Since they are clearly not processes or compositions, they fail to fall within a statutory category. 2) The media must be structurally and functionally  
20 interconnected with the functional descriptive material in such a manner that it enables the functional descriptive material to act as a computer component and realize its functionality. The current office position is that signals, **waves**, radiation, wires, fibers, and **printed matter** in and of themselves fail to meet this criteria. As such, no usefulness can be gleaned from the functional descriptive material. 3) The functional descriptive material must provide a practical application of the idea embodied therein when executed as a computer component. Following  
25 Office policy, Examiner maintains that claim 71 is directed to non-statutory subject matter for at least these reasons. On this basis, Examiner finds Applicant's argument to be unpersuasive and the rejections STAND.

In re pages 12-13, Applicant argues that the claim limitations are directed to the method actually used by the invention. Examiner clarifies: The intent of the quoted sentence from page 4 of the Office Action mailed 11/2/2006

Art Unit: 2129

was: "no claim limitations are directed to the method of measuring closeness actually used by the invention." This has been corrected in the current Office Action as seen above.

In re page 13, Applicant argues that preemption rejection. Examiner finds these arguments persuasive.

Preemption is no longer a basis for the rejection under 35 U.S.C. §101 detailed above.

5 Examiner finds that the claims are not statutory. The rejection of claims 53-59, 65, and 71 under 35 U.S.C. §101 is **MAINTAINED**.

### **Claim Rejections - 35 USC §103**

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this

10 Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

15

Claims 53-59, 65, and 71 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Liddy** (USPN 5,963,940) and **Wermter** ("Recurrent Neural Network Learning for Text Routing").

20 **Claims 53, 65, and 71:**

**Liddy** teaches:

- wherein each dataset includes one or more data points and each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string (C2:37 especially "texts ... proper nouns (PNs), single terms, text structure ... document database ... words ... documents ... text structure ... intentions ... tables, graphs, photographs or other images ... captions" C2:48-25 C3:41);
- constructing a semantic vector for each dataset (C2:37 especially "documents in the database have preferably been processed to provide corresponding alternative representations" C2:60-65 or "semantic" C6:60-C7:7 or "Natural Language Processing" or "subject-based vector representation of the document's 30 contents" C9:30-60; Also see the Abstract);
- receiving a query containing information indicative of desired datasets (C2:37 especially "query" C2:45-30 C3:20 or C12:50-60; Also see the Abstract);

Art Unit: 2129

- constructing a semantic vector for the query (C2-37 especially "processes the query to generate an alternative representation ... is a subject field code vector" C2:45-67 or "texts (document and queries) are processed to determine discourse aspects of the text" C3:20-35 or "semantic" C6:60-C7:7 or "Natural Language Processing" or "subject-based vector representation of the document's contents" C9:30-60 or "Query Processing" C16:48-C17:10; Also see the Abstract);
- comparing the semantic vector for the query to the semantic vector of each dataset (C2-37 especially "The matching score between the query and document is determined ... computes the distance between these two data points" C23:20-35 or "query representation is matched to the relevant document database" C2:55-65); and
- selecting datasets whose semantic vectors are closest in distance to the semantic vector for the query (C2-37 especially "The matching score between the query and document is determined ... computes the distance between these two data points" C23:20-35 or "displays documents judged relevant to the content of the query" C3:40-60).
- outputting information of the selected datasets to be corresponding to the desired datasets identified in the query (C2-37 especially "displays documents judged relevant to the content of the query" C3:40-64);
- wherein:
  - o the query or each of the datasets includes at least one data point (p898-903 especially ); and
  - o the semantic vector for the query or each of the datasets is constructed by the steps of:
    - for each data point, identifying a relationship between each data point and categories in the semantic space (C2-37 especially "Each information bearing word in a text is looked up in the online, lexical resource ... representation of the document's contents" C9:50-60);
    - determining the significance of each data point with respect to the categories, wherein the significance represents a relative strength of each data point relative to each of the categories, or a degree of relevance of each data point relative each of the categories (C2-37 especially "Each information bearing word in a text is looked up in the online, lexical resource ... representation of the document's contents" C9:50-60);

Art Unit: 2129

- constructing a semantic vector for each data point, wherein each semantic vector represents the relative strength of its corresponding data point with respect to each category (C2-37 especially "vector representation" C9:30-60); and
- based on the semantic vector for each of the at least one data point, form the semantic vector of the query or each of the datasets (C2-37 especially "The vector for the original query and ... documents are weighted and combined to form a new, single vector for re-ranking and re-Clustering" C26:10-17 and "vector representation" C9:30-60).

**Liddy** fails to teach:

- predetermined categories corresponding to dimensions in the semantic space; and
- wherein each semantic vector has dimensions equal to the number of predetermined categories.

**Wermter** teaches:

- wherein each dataset includes one or more data points and each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string (p898-903 especially "text categorization test collection [9] contains real-world documents" §5 or "words in the corpus" §5);
- constructing a semantic vector for each dataset (p898-903 especially "semantic significance vectors" §5);
- wherein:
  - the query or each of the datasets includes at least one data point (p898-903 especially "text categorization test collection [9] contains real-world documents" §5 or "words in the corpus" §5); and
  - the semantic vector for the query or each of the datasets is constructed by the steps of:
    - for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space (p898-903 especially "belong to one of eight main categories ... the 8 separate categories ... words in the corpus are represented using semantic significance vectors ... where  $c_i$  represents a certain semantic category. A value  $v(w, c_i)$  is computed for each dimension of the semantic vector" §5);
    - determining the significance of each data point with respect to the predetermined categories, wherein the significance represents a relative strength of each data point relative to each of the predetermined particular categories, or a degree of relevance of each data point relative

Art Unit: 2129

each of the predetermined particular categories (p898-903 especially "belong to one of eight main categories ... the 8 separate categories ... words in the corpus are represented using semantic significance vectors ... where  $c_i$  represents a certain semantic category. A value  $v(w, c_i)$  is computed for each dimension of the semantic vector" §5);

- constructing a semantic vector for each data point, wherein each semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories (p898-903 especially "belong to one of eight main categories ... the 8 separate categories ... words in the corpus are represented using semantic significance vectors ... where  $c_i$  represents a certain semantic category. A value  $v(w, c_i)$  is computed for each dimension of the semantic vector" §5); and
- based on the semantic vector for each of the at least one data point, form the semantic vector of the query or each of the datasets (p898-903 especially "belong to one of eight main categories ... the 8 separate categories ... words in the corpus are represented using semantic significance vectors ... where  $c_i$  represents a certain semantic category. A value  $v(w, c_i)$  is computed for each dimension of the semantic vector" §5).

#### Motivation:

**Liddy** and **Wermter** are from the same field of endeavor, text processing. It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the teachings of **Liddy** by using the semantic vectors having dimensions equal to a predetermined number of categories as taught by **Wermter** for the benefit of encoding the sequential context of word sequences reaching high recall and precision rates (**Wermter** §1).

#### **Claim 54:**

**Liddy** anticipates:

- Wherein the datasets correspond to documents and the query is a natural language query (C2:37 especially "natural language query" C2:48-C3:15 and C5:1-7).

Art Unit: 2129

**Claim 55:****Liddy** anticipates:

- Performing a second search for datasets within the collection of datasets, wherein the second search using a method other than semantic vectors (C2-37 especially "proper noun matching" C24:64-25:15 and "Complex Nominals ... Mandatory requirements" C21:25-55);
- Combining the two search results to obtain a combined weighted score for each dataset in either of the two search results (C2-37 especially "Weighted Boolean Processor" C20:54-C21:20 and "Scoring" C22:1-8);
- Selecting datasets whose combined weighted score is largest (C2-37 especially "highest weighted score" C21:1-13).

**Claim 56:****Liddy** anticipates:

- Further comprising a step of clustering the selected datasets in real time (C2-37 especially "real time" C7:20-35).

**Claim 57:****Liddy** anticipates:

- wherein each dataset includes one or more data points and each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string (C2-37 especially "texts ... proper nouns (PNs), single terms, text structure ... document database ... words ... documents ... text structure ... intentions ... tables, graphs, photographs or other images ... captions" C2:48-C3:41);
- constructing a semantic vector for the dataset (C2-37 especially "documents in the database have preferably been processed to provide corresponding alternative representations" C2:60-65 or "semantic" C6:60-C7:7 or "Natural Language Processing" or "subject-based vector representation of the document's contents" C9:30-60; Also see the Abstract);

Art Unit: 2129

- comparing the semantic vector for the dataset to a semantic vector of each of the data points in the semantic lexicon (C2-37 especially "The matching score between the query and document is determined ... computes the distance between these two data points" C23:20-35 or "query representation is matched to the relevant document database" C2:55-65 or "lexicon" C9:50-60 or C17:15-35);
- 5 - selecting data points whose trainable semantic vectors are closest in distance to the semantic vector for the dataset (C2-37 especially "The matching score between the query and document is determined ... computes the distance between these two data points" C23:20-35 and "displays documents judged relevant to the content of the query" C3:40-60); and
- associating said selected data points to said dataset (C2-37 especially "The vector for the original query and  
10 ... documents are weighted and combined to form a new, single vector for re-ranking and re-Clustering" C26:10-17);
- wherein:
  - o the semantic vector for the dataset is constructed by the steps of:
    - for each data point, identifying a relationship between each data point and categories in the  
15 semantic space (C2-37 especially "Each information bearing word in a text is looked up in the online, lexical resource ... representation of the document's contents" C9:50-60);
    - determining the significance of each data point with respect to the categories, wherein the significance represents a relative strength of each data point relative to each of the categories, or a degree of relevance of each data point relative each of the categories (C2-  
20 37 especially "Each information bearing word in a text is looked up in the online, lexical resource ... representation of the document's contents" C9:50-60);
    - constructing a semantic vector for each data point, wherein each semantic vector represents the relative strength of its corresponding data point with respect to each category (C2-37 especially "vector representation" C9:30-60); and
    - 25 ▪ based on the semantic vector for each of the at least one data point, form the semantic vector of the dataset (C2-37 especially "The vector for the original query and ... documents are weighted and combined to form a new, single vector for re-ranking and re-Clustering" C26:10-17 and "vector representation" C9:30-60).



Art Unit: 2129

**Liddy** fails to teach:

- predetermined categories corresponding to dimensions in the semantic space; and
- wherein each semantic vector has dimensions equal to the number of predetermined categories.

**Wermter** teaches:

- 5        - wherein each dataset includes one or more data points and each data point corresponds to at least one of a word, a phrase, a sentence, a color, a typography, a punctuation, a picture, and a character string (p898-903 especially "text categorization test collection [9] contains real-world documents" §5 or "words in the corpus" §5);
- constructing a semantic vector for each dataset (p898-903 especially "semantic significance vectors" §5);
- 10       - wherein:
  - o the query or each of the datasets includes at least one data point (p898-903 especially "text categorization test collection [9] contains real-world documents" §5 or "words in the corpus" §5); and
  - o the semantic vector for the dataset is constructed by the steps of:
    - for each data point, identifying a relationship between each data point and predetermined
    - 15                categories corresponding to dimensions in the semantic space (p898-903 especially "belong to one of eight main categories ... the 8 separate categories ... words in the corpus are represented using semantic significance vectors ... where  $c_i$  represents a certain semantic category. A value  $v(w, c_i)$  is computed for each dimension of the semantic vector" §5);
    - determining the significance of each data point with respect to the predetermined categories,
    - 20                wherein the significance represents a relative strength of each data point relative to each of the predetermined particular categories, or a degree of relevance of each data point relative each of the predetermined particular categories (p898-903 especially "belong to one of eight main categories ... the 8 separate categories ... words in the corpus are represented using semantic significance vectors ... where  $c_i$  represents a certain semantic category. A value
    - 25                 $v(w, c_i)$  is computed for each dimension of the semantic vector" §5);
    - constructing a semantic vector for each data point, wherein each semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories

Art Unit: 2129

(p898-903 especially "belong to one of eight main categories ... the 8 separate categories ... words in the corpus are represented using semantic significance vectors ... where  $c_i$  represents a certain semantic category. A value  $v(w, c_i)$  is computed for each dimension of the semantic vector" §5); and

- 5                   ▪ based on the semantic vector for each of the at least one data point, form the semantic vector of the dataset (p898-903 especially "belong to one of eight main categories ... the 8 separate categories ... words in the corpus are represented using semantic significance vectors ... where  $c_i$  represents a certain semantic category. A value  $v(w, c_i)$  is computed for each dimension of the semantic vector" §5).

10    Motivation:

**Liddy** and **Wermter** are from the same field of endeavor, text processing. It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the teachings of **Liddy** by using the semantic vectors having dimensions equal to a predetermined number of categories as taught by **Wermter** for the benefit of encoding the sequential context of word sequences reaching high recall and precision rates

15    (**Wermter** §1).

**Claim 58:**

**Liddy** anticipates:

- 20           - Wherein the dataset is a document and the data points are words (C2-37 especially "all words in the document" C9:50-60 and "documents ... words" C15:24-32 and "word" and "document" C16:1-25).

**Claim 59:**

**Liddy** anticipates:

- 25           - Wherein the dataset is a natural language query in a search system and the data points are words (C2-37 especially "natural language ... words" C15:24-32).

**Conclusion**

Claims 53-59, 65, & 71 are rejected.

Art Unit: 2129

**Correspondence Information**

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Benjamin Buss whose telephone number is 571-272-5831. The examiner can normally be reached on M-F 9AM-5PM.

As detailed in MPEP 502.03, communications via Internet e-mail are at the discretion of the applicant. Without a written authorization by applicant in place, the USPTO will not respond via Internet e-mail to any Internet correspondence which contains information subject to the confidentiality requirement as set forth in 35 U.S.C. 122. A paper copy of such correspondence will be placed in the appropriate patent application. The following is a sample authorization form which may be used by applicant:

"Recognizing that Internet communications are not secure, I hereby authorize the USPTO to communicate with me concerning any subject matter of this application by electronic mail. I understand that a copy of these communications will be made of record in the application file."

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David Vincent can be reached on 571-272-3080. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Benjamin Buss  
Examiner  
Art Unit 2129

  
DAVID VINCENT  
SUPERVISORY PATENT EXAMINER